2

ADA124838

COMPUTER SCIENCE TECHNICAL REPORT SERIES

DTIC
ELECTE
FEB 24 1983
S D
B

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND
20742

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

DTIC FILE COPY

83 02 023 133

TR-1188
DAAG-53-76C-0138
August 1982

MOTION ESTIMATION BASED ON MULTIPLE LOCAL CONSTRAINTS AND NONLINEAR SMOOTHING

Kwangyoen Wohn
Larry S. Davis
Philip Thrift

Computer Vision Laboratory
Computer Science Center
University of Maryland
College Park, MD 20742

ABSTRACT

A new algorithm for estimating motion from image sequences is presented. Initial motion estimates are determined based on a least-squares solution to a set of independent linear constraints on the motion at a pixel. These initial estimates are then improved by a nonlinear smoothing operation. The results of this algorithm are compared with those obtained by the Horn-Schunck algorithm ~~[10]~~ 9 on a number of image sequences.

DTIC
S ELECTE D
FEB 24 1983
B

The support of the Defense Advanced Research Projects Agency and the U.S. Army Night Vision Laboratory under Contract DAAG-53-76C-0138 (DARPA Order 3206) is gratefully acknowledged, as is the help of Janet Salzman in preparing this paper.

*Current address: Texas Instruments, Inc., Dallas, TX

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

1. Introduction

One of the key problems in analysis of time-varying images is the computation of object motion from frame to frame. Huang [1] identifies three approaches to motion estimation from image sequences:

1) Fourier methods

2) Correspondence methods

3) Differential methods.

Fourier methods are based on the observation that a 2-D translation and/or rotation of an image has a simple effect on the Fourier transform of the image. For example, a translation results in a phase shift of the transform, so that differences in the appropriate phase angles from frame to frame determine a translation. For this approach to work, the object must be translating across a homogeneous ("flat") background. However, extensions to rotation (and scale) involve more complicated comparison of transforms, which makes the approach computationally unattractive.

Correspondence methods are two stage methods and involve

1) finding image points in successive frames which correspond to the projection of the same scene point, and

2) using those correspondences to solve for the motion.

Step 2 is straightforward in principle, although it may involve non-linear parameter estimation (see Ullman [2], Huang [1] and Huang and Tsai [3]). If the motion changes slowly with time, then this is not a severe practical problem.

For step 1, we must determine reliably identifiable image features which are invariant to motion. For example, if image boundaries correspond to locally planar reflectance contours on 3-D objects, then curvature discontinuities and curvature zero-crossings are features invariant to motion, and can be used to compute the correspondences in step 1. For a survey of matching algorithms for solving the correspondence problem, see [4].

Differential methods are based on the relationship between motion and spatial/temporal image intensity derivatives, namely that

$$-I_t = I_x u + I_y v \tag{1}$$

where I_t is the temporal intensity derivative, I_x and I_y are the spatial intensity derivatives in the x and y directions, and u and v are the components of the motion in the x and y directions. Equation (1) can be derived using a Taylor series expansion for the image function and assuming a) that gray level is invariant to motion and b) that the image intensity function is locally planar, i.e., higher order spatial intensity derivatives are zero. Equation (1) represents a linear constraint on u and v, so that while the component of motion in the gradient direction can be determined (the "normal component"), the component in the level direction cannot. However, by combining the motion constraints from a number of points, one can compute an estimate of the image motion.

These constraints are combined based on assumptions about the image vectors. For example, in [5,6] the assumption is made that the image motion vectors are locally constant (corresponding to an image plane translation) so that the least-square pseudo-intersection of the constraint lines in (u,v) space represents the motion vector at each point. Since such approaches require combining the normal velocity components of neighboring points, they will not yield reliable estimates near the boundaries of moving objects. In Section 2 of this paper we show how multiple linear constraints on image motion can be derived at individual points so that a motion estimate can be computed for a point without making any assumptions about the spatial pattern of motion vectors.

Unfortunately, the motion vectors computed using this approach (or the pseudo-intersection approach) are not reliable, so that it is desirable to perform some (perhaps iterative) spatial smoothing of the vectors. An important property of such smoothing schemes is that they should not smooth across the boundary of a moving object. In Section 3 of this paper we present such a technique. We assume that the image motion is locally a rigid 2½D motion (see also Schalkoff and McVey [7]), and smooth the velocity vector at a point based on associating it with that one of its 8-neighbors whose neighborhood is "most smooth" in this sense (compare Haralick [8]). For interior points near the boundary of a moving object, we could expect such an approach to choose a neighborhood which is completely contained within the object.

2. Multiple constraint equations of a single point

The differential methods discussed in Section 1 are based on the assumption that the image intensity corresponding to any scene point is invariant to motion; i.e., if $I(x,y,t)$ is the intensity at position (x,y) and time t, and if the image motion is (u,v), then $I(x,y,t)=I(x+u,y+v,t+1)$. Since the intensity is invariant to motion, so are various derivatives of intensity. In particular, the gradient of intensity would be invariant to motion, so that we can write a second linear constraint on the image motion:

$$G_x u + G_y v = -G_t \tag{2}$$

where G_x and G_y are the spatial derivatives of the image gradient (i.e., directional second derivatives) and G_t is the temporal derivative of the gradient.

In general, if F denotes any motion-invariant feature, we can produce the constraint equation

$$F_x u + F_y v = -F_t \tag{3}$$

For example, F may be the gradient direction, the curvature of the surface, the moments of local intensity distributions, or higher-order derivatives of intensity. However, in practice, features that are defined in terms of higher-order spatial derivatives are not reliable since the differentiation operation tends to amplify noise.

Notice that we can also construct sets of constraint equations from color images, since at least one constraint equation may be obtained from each band. If more than two constraint equations are available at any point, then a least-square solution to the pseudo-intersection of these constraints can be obtained using pseudo-inverse techniques.

DTIC
COPY
INSPECTED
2

Accession For
NTIS GRA&I ☑
DTIC TAB □
Unannounced □
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A

3. Iterative algorithm for motion enhancement

The velocity vectors computed either by spatially integrating constraints from local neighborhoods of points or combining multiple constraint equations from a single point are often inaccurate. In this section we introduce an iterative algorithm for enhancing such motion vectors.

The method for smoothing motion vectors is based on the following two assumptions:

1. Objects are rigid, and
2. Objects are undergoing 2½D motion, i.e., they move along a plane perpendicular to the line of sight with arbitrary translational and rotational velocity.

We iteratively update the velocity at point P on the image as follows. First, choose the neighbor of P whose 3×3 neighborhood of motion vectors best fits a 2½D rigid motion. Let the rotational velocity in that neighborhood be ω. Since the rotational velocity is constant over a single moving object, we regard ω as the rotational velocity for P and update the velocity at P. The details are described below.

Figure 1 shows an object moving with angular velocity $\vec{\Omega}$ and translational velocity v^*_T. O* is the center of rotation. For an arbitrary point P on the object surface, the resultant velocity $\vec{v}$ is

$$\vec{v} = \vec{v}_T + \vec{r} \times \vec{\Omega} \tag{4}$$

Given the image motion vectors at P and P', we compute $\vec{\omega}$ (the rotational velocity of P' about P) as follows:

$$\Delta\vec{v} = \vec{v}' - \vec{v} \tag{5}$$

$$\Delta\vec{r} = \vec{R}' - \vec{R} = \vec{r}' - \vec{r} \tag{6}$$

The rotational velocity $\vec{\omega}'$ at P' with respect to P is then

$$\vec{\omega}' = \frac{\Delta\vec{r} \times \Delta\vec{r}}{\|\Delta\vec{r}\|^2} \tag{7}$$

The rotational velocity at P is obtained by averaging $\vec{\omega}'$ for all points in a neighborhood, N, of P of size n:

$$\vec{\omega} = \sum_{i \in N} \vec{\omega}^i/n \tag{8}$$

To compute motion along the line of sight, we consider the quantity

$$D = \sum_{i \in N} \frac{\Delta\vec{r} \cdot \Delta\vec{v}}{\|\Delta\vec{r}\|^2} \tag{9}$$

D is a measure of dilation, which reflects motion along the line of sight. For 2½D object motion D should be constant for all points in a moving object. (D>0 if the relative depth is decreasing.)

Next we consider the following two error measurements at an arbitrary point i:

$$E_i^1 = \sum_{j \in N} \|\vec{\omega}_j - \vec{\omega}_i\|^2 \tag{10}$$

$$E_i^2 = \sum_{j \in N} [D_j - D_i] \tag{11}$$

To update the velocity at point P, we choose a point P' from P's neighborhood such that the linear combination $E_{P'}$ of the above two errors is a minimum at P':

$$E_{P'} = \min_{j \in N} [E_j^1 + \alpha \cdot E_j^2] \tag{12}$$

where α is a scalar. The velocity $\vec{v}$ at P is then adjusted by assuming that P has the rotational velocity $\vec{\omega}'$ with respect to P' and computing

$$\vec{v} = \vec{v}' - \Delta\vec{r} \times \vec{\omega}' \tag{13}$$

where $\vec{v}'$ is the translational velocity at P' and $\Delta\vec{r}$ is the vector joining P and P'. An important advantage of this approach is that since the error measurements along a moving boundary are relatively large, the enhancement tends not to combine the motion estimates across such boundaries.

4. Experiments

In this section we describe a set of experiments designed to demonstrate the behavior of the motion estimation and smoothing algorithms described in the previous sections. For comparison, we also implemented one other motion estimation algorithm (the pseudo-intersection method mentioned in Section 1) and one of the motion smoothing algorithms (the one described in Horn and Schunck [9] - although note that in [9] the original motion vector field was just the field of normal components).

The input image sequences are displayed in Figures 2-4. Figure 2 shows a sequence that contains two moving cars. Figure 3 and 4 are synthetic images. In Figure 3, a sphere rotates in the image plane, while approaching the viewer (a 2-D rotation with zoom). In Figure 4, the same sphere undergoes a 3-D rotation, while approaching the viewer and translating towards the right. The intensity corresponding to any point on the sphere is invariant to the motion.

Although, in principle, it is possible to use the multiple constraint method to compute a velocity vector at each point, in fact a number of practical considerations limit the set of points at which useful estimates can be obtained to points:

1) having non-zero spatial and temporal intensity derivatives,
2) having non-singular matrices corresponding to the multiple constraint equations, and
3) for which the estimated motion is small (i.e., vector magnitude less than 5 pixels).

Figure 5a shows the motion estimates for the multi-constraint method for frame 6 of the moving car sequence. All spatial derivatives are based on fitting a quadratic surface to a 5×5 neighborhood, and all temporal gradients are based on a quadratic approximation to 5 consecutive temporal points. For comparison, Figure 5b shows the motion vectors computed by the pseudo-intersection method. They are much better approximations to the actual motion, although as we shall see, the motion enhancement algorithm plays a larger role than the initial estimates in determining the accuracy of the final motion vector field.

Next, we consider three combinations of initial motion estimation and motion enhancement:

A) Multi-constraint initial estimation with Horn and Schunck motion enhancement.

B) Pseudo-intersection initial estimation with Horn and Schunck motion enhancement.

C) Pseudo-intersection initial estimation with the motion enhancement algorithm of Section 3.

By comparing A and B we can evaluate the role of the initial estimates in the overall motion computation, while the comparison of B with C can demonstrate whether the computationally more costly algorithm of Section 3 (which is designed to avoid smoothing over motion boundaries) has its higher cost justified by better motion estimates.

In order to evaluate the results quantitatively, we consider two error measures. The first measure was designed to enable us to evaluate motion estimates on sequences for which the actual motion is unknown and is based on measuring how well the motion vectors predict intensity from one frame to the next. This measure, E_p, is defined as

$$E_p(t) = \frac{1}{n} \sum_{i \in L} |I^{t+1}_{i+v_i} - I^t_i|$$

where:

a) I^t_i is the intensity of pixel i at time t,

b) L is the set of pixels at which the estimated motion vector is non-zero, and

c) n is the size of L.

The second measure, E_v, compares the estimated motion vectors with true motion vectors, and is defined as

$$E_v(t) = \frac{1}{m} \sum_{i \in k} \|\vec{v}^*_{i,t} - \vec{v}_{i,t}\|$$

where

a) $\vec{v}^*_{i,t}$ is the true motion vector at point i at time t,

b) $\vec{v}_{i,t}$ is the estimated motion vector at point i at time t,

c) K is the set of points having non-zero true motion, and

d) m is the size of the set K.

Figure 6a-d show E_p for four frames of the car sequence in Figure 2. We can make the following observations about these graphs (the observations also hold for other frames in this sequence):

1) The final motion vectors are not particularly sensitive to the choice of initial estimate, but appear to depend more critically on the enhancement algorithm. (Compare A-B.)

2) Most of the enhancement takes place during the first two iterations of either enhancement algorithm.

3) After 3 iterations the differences between all three approaches are insignificant, so that we would choose among them based on computational cost (which would lead to a choice of B - pseudo-intersection with Horn and Schunck.) Since the multi-constraint method offers no practical advantages over pseudo-intersection method for initial motion estimation, we will not consider method A in the remaining examples.

Consider next the spheres in Figures 3 and 4. Figures 7 and 8 show E_v for one frame from each of these sequences.

Here, curve A corresponds to the non-linear enhancement algorithm in Section 3 and curve B corresponds to the Horn-Schunck algorithm. Furthermore, we have decomposed the total error into two components - one corresponding to a region near the border of the sphere (A-1 and B-1), and the second corresponding to the interior of the sphere (A-2, B-2). We observe that for both the 2-D and the 3-D motion, the error component due to the boundary is less for the non-linear enhancement algorithm than for the Horn-Schunck algorithm. Thus, near the boundary, at least, the search for a "best" neighborhood to compute the enhancement does lead to more accurate motion estimates. Considering the component of error on the interior of the sphere, we note that both the non-linear algorithm and Horn-Schunck produce very accurate motion estimates for the 2½D motion (errors of ~.1). This is not surprising, since the non-linear algorithm is explicitly based on a 2½D motion assumption, while as regards Horn-Schunck, the Laplacian of a 2½D motion vector field is zero. The two algorithms produce similar, but higher, errors in the 3-D motion case, the slightly better performance of the nonlinear algorithm perhaps attributable to the search for a best neighborhood.

5. Conclusions

Based on the experiments presented in Section 4 we can draw the following conclusions:

1. The multi-constraint algorithm does not produce reliable motion estimates.
2. The enhancement algorithm plays a much larger role than the initial estimates in determining the utility of the final motion estimates.
3. Enhancing the motion estimate of a pixel based on first searching for the "best" neighborhood containing that pixel (i.e., the neighborhood whose motion estimates best satisfy the given motion model) yields much more accurate motion estimates near the borders of moving regions.

References

1. T. S. Huang and R. Y. Tsai, "Image Sequence Analysis: Motion Estimation", in Image Sequence Analysis (ed. T. S. Huang), Springer-Verlag, Berlin, 1981, 1-18.

2. S. Ullman, The Interpretation of Visual Motion, MIT Press, Cambridge, MA, 1979.

3. R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces," Proc. PRIP '82, Las Vegas, NV, June 14-17, 1982, 112-118.

4. J. K. Aggarwal, L. S. Davis, and W. N. Martin, "Correspondence processes in dynamic scene analysis," Proc. IEEE 69, 1981, 562-572.

5. C. Cafforio and F. Rocca, "Tracking moving objects in TV images." Signal Processing, 1, 1979, 133-140.

6. J. Limb and J. Murphy, "Estimating the velocity of moving images in TV signals," Computer Graphics and Image Processing, 4, 1975, 311-327.

7. R. J. Schalkoff and E. S. McVey, "A model and tracking algorithm for a class of video targets," IEEE Trans. PAMI, 4, 1982, 2-10.

8. R. M. Haralick and L. Watson, "A facet model for image data," Computer Graphics and Image Processing, 15, 1981, 113-129.

9. B. K. P. Horn and B. G. Schunck, "Determining Optical Flow," Artificial Intelligence, 17, 1981, 185-204.

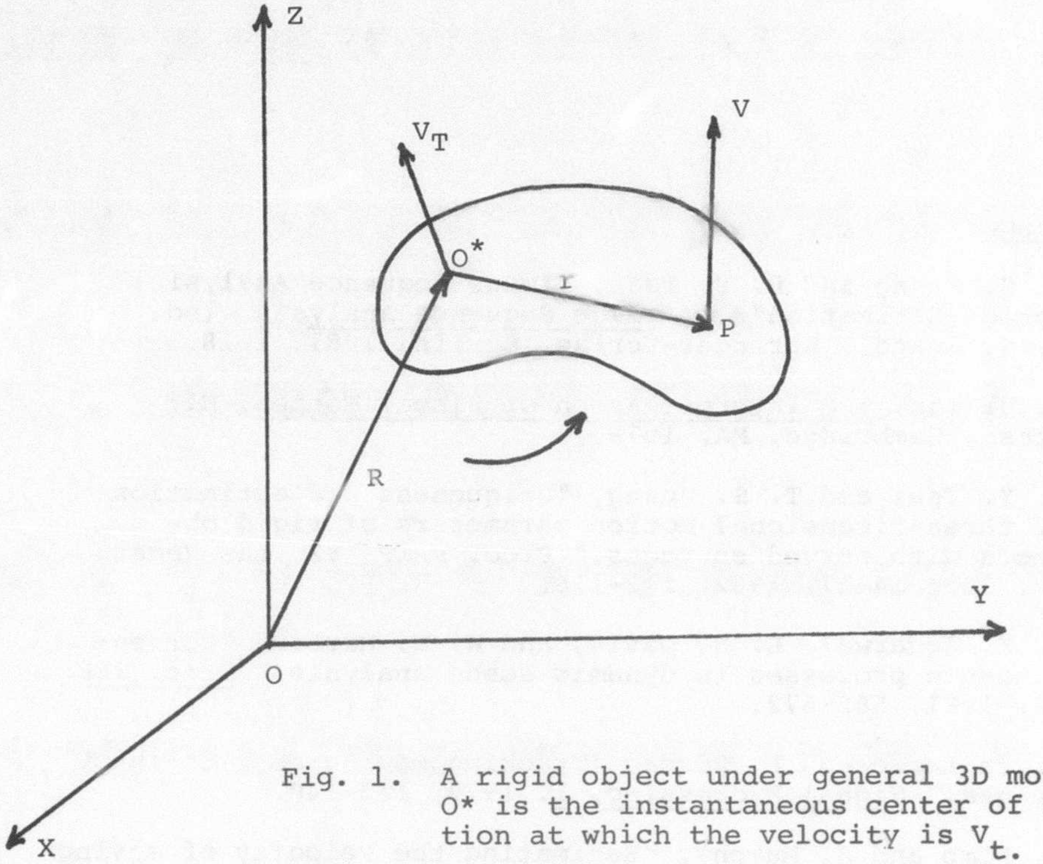

Fig. 1. A rigid object under general 3D motion. O* is the instantaneous center of rotation at which the velocity is V_t.

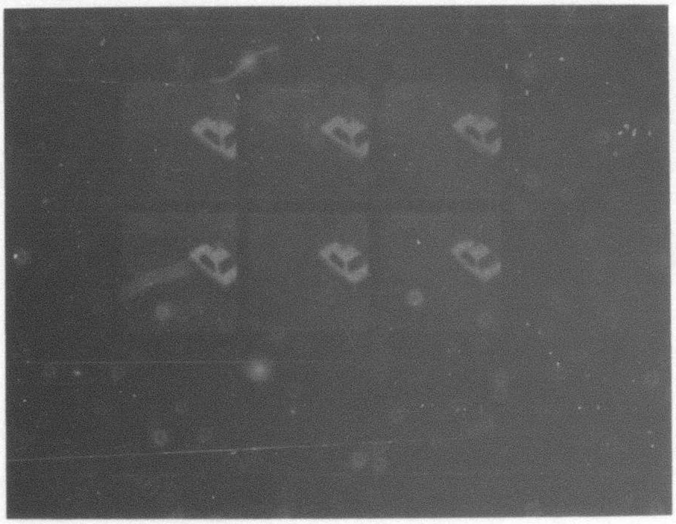

Fig. 2. Moving cars.

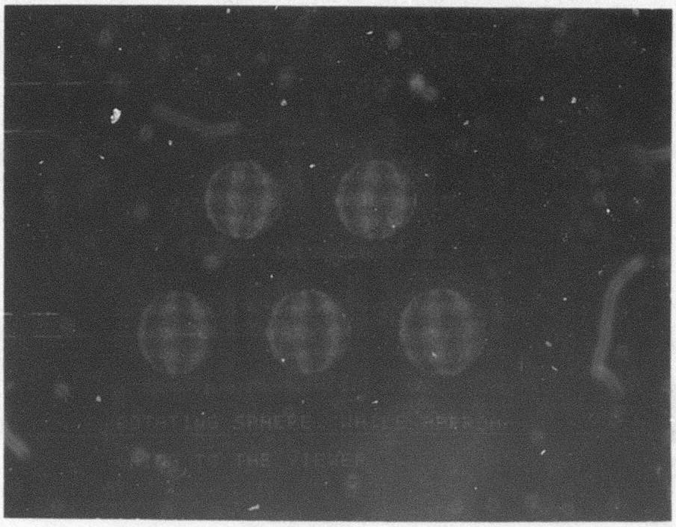

Fig. 3. Sphere undergoing a 2½D motion.

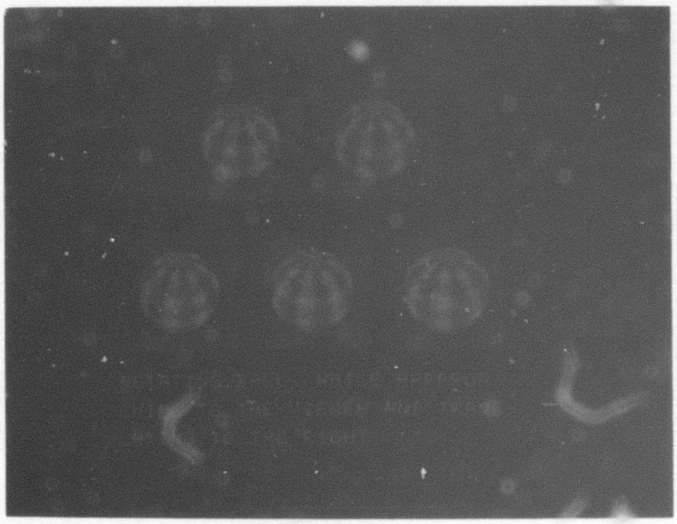

Fig. 4. Sphere undergoing a 3-D motion.

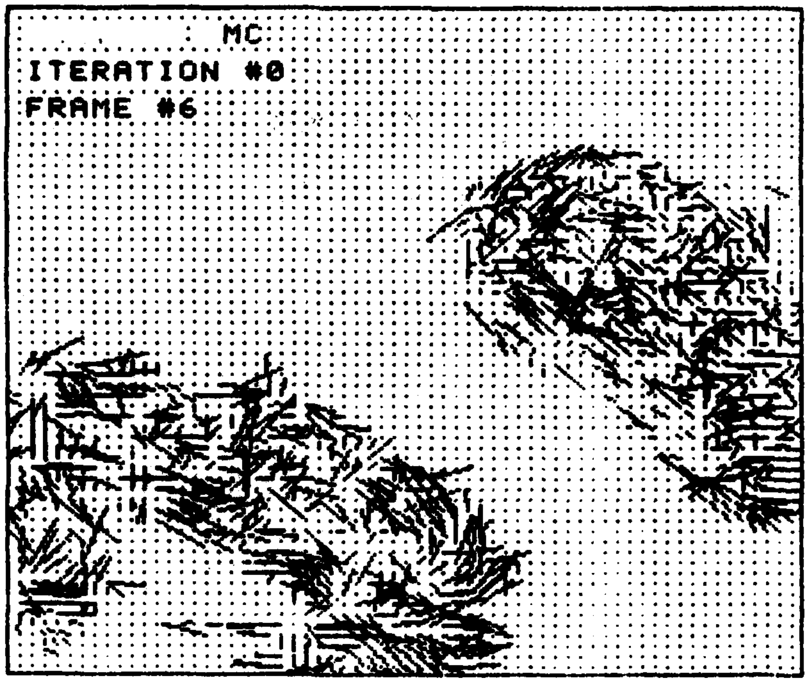

Fig. 5a. Initial estimates using multi-constraints.

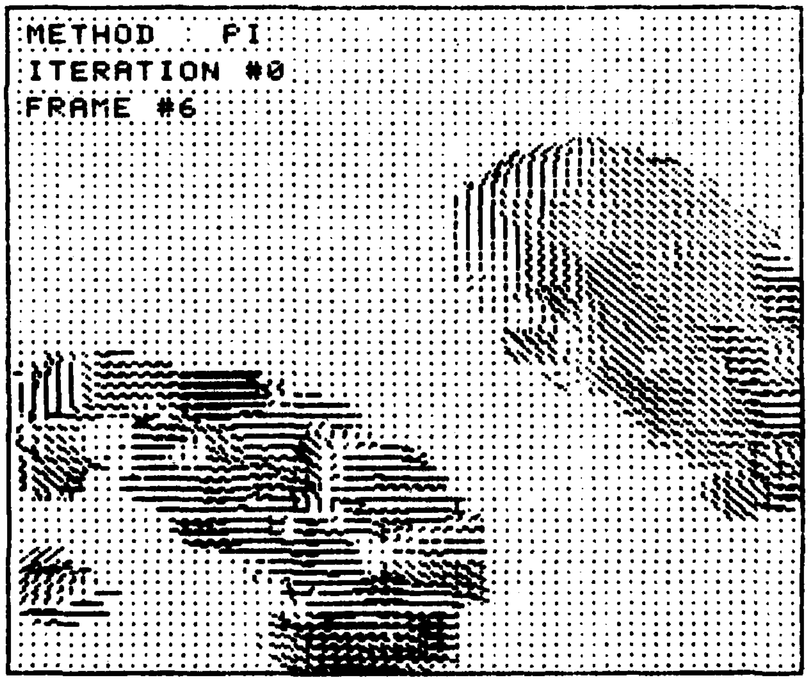

Fig. 5b. Initial estimates using pseudo-intersection.

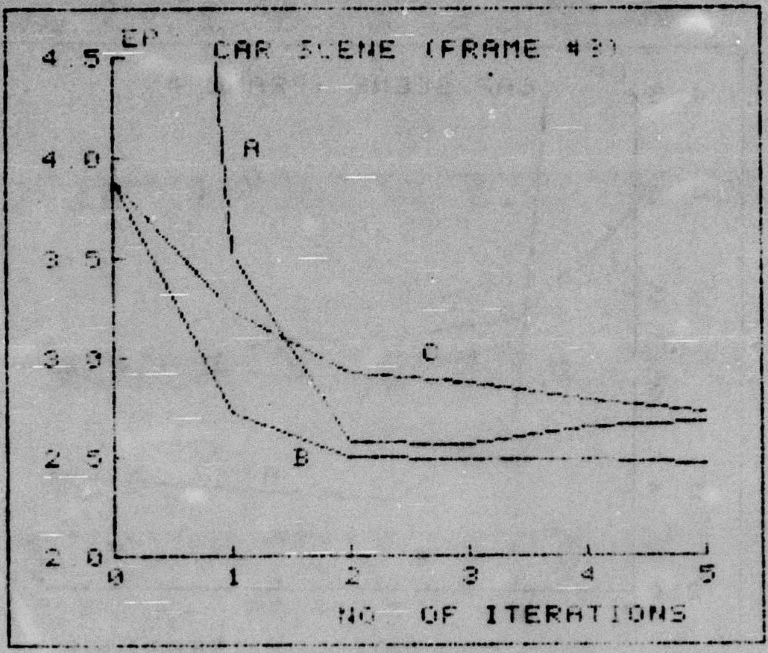

Fig. 6a. A - Multi-constraint with Horn-Schunck.
B - Pseudo-intersection with HS.
C - PI with facet-like enhancement of Section 3.

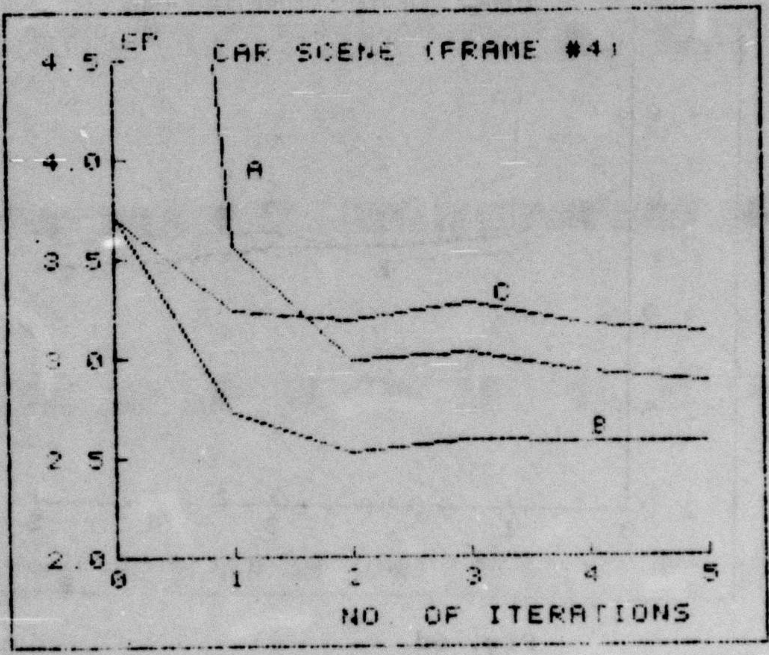

Fig. 6b.

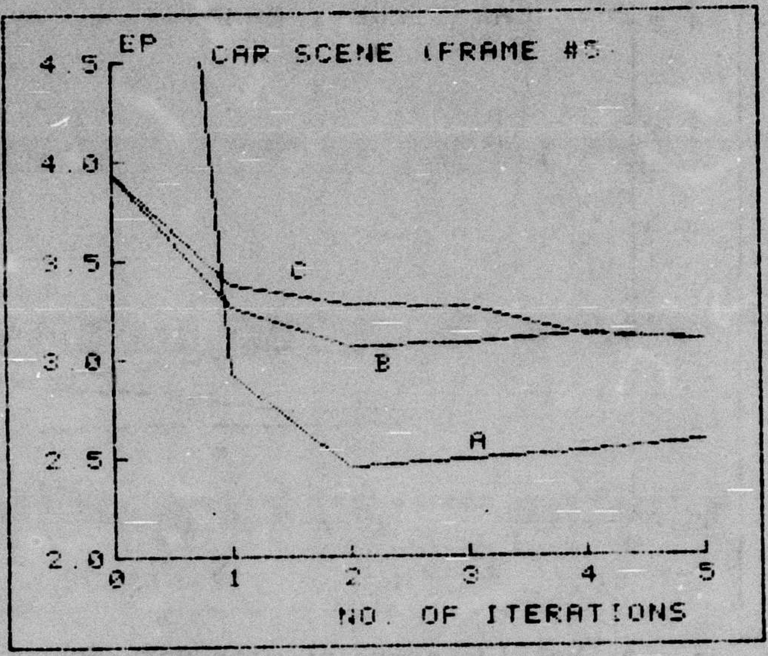
EP
CAR SCENE (FRAME #5
4.5
4.0
3.5
3.0
2.5
2.0
0
1
2
3
4
5
C
B
A
NO. OF ITERATIONS

Fig. 6c.

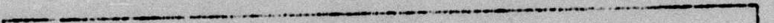

Fig. 6d.

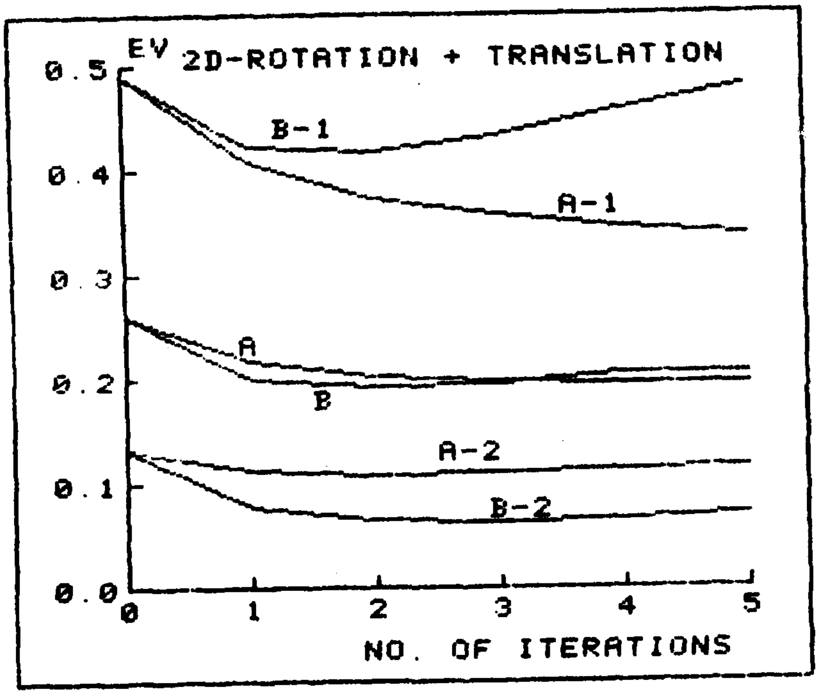
EV
2D-ROTATION + TRANSLATION
0.5
0.4
0.3
0.2
0.1
0.0
B-1
A-1
A
B
A-2
B-2
0
1
2
3
4
5
NO. OF ITERATIONS

Fig. 7.

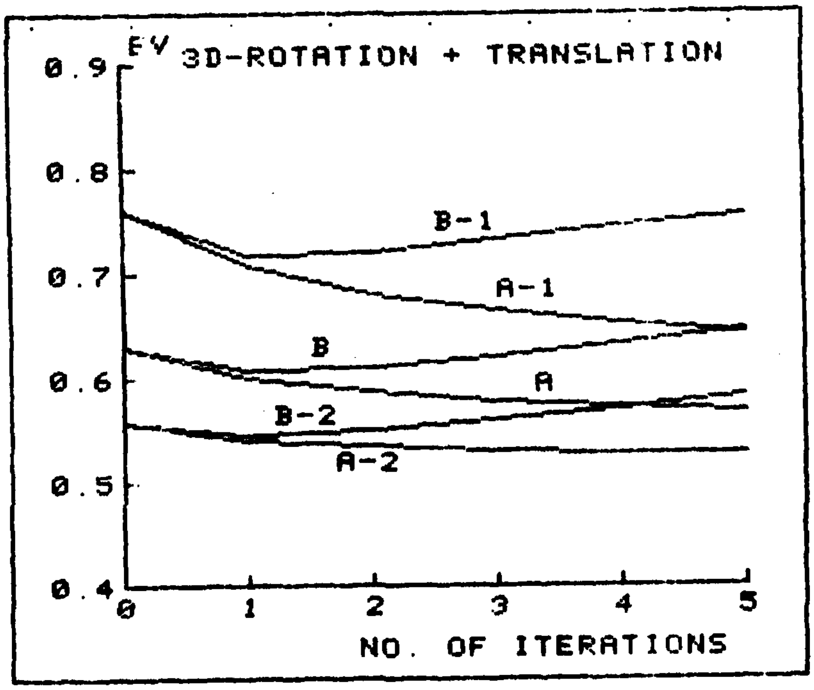
EV
3D-ROTATION + TRANSLATION
0.9
0.8
0.7
0.6
0.5
0.4
B-1
A-1
B
A
B-2
A-2
0
1
2
3
4
5
NO. OF ITERATIONS

Fig. 8.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO. AD-A124 838

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
MOTION ESTIMATION BASED ON MULTIPLE LOCAL CONSTRAINTS AND NONLINEAR SMOOTHING

5. TYPE OF REPORT & PERIOD COVERED
Technical

6. PERFORMING ORG. REPORT NUMBER
TR-1188

7. AUTHOR(s)

Kwangyoen Wohn
Larry S. Davis
Philip Thrift

8. CONTRACT OR GRANT NUMBER(s)
DAAG-53-76C-0138

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Computer Vision Laboratory
Computer Science Center
University of Maryland
College Park, MD 20742

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
U.S. Army Night Vision Laboratory
Ft. Belvoir, VA 22060

12. REPORT DATE
August 1982

13. NUMBER OF PAGES
23

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Image processing
Time-varying imagery
Motion estimation
Smoothing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A new algorithm for estimating motion from image sequences is presented. Initial motion estimates are determined based on a least-squares solution to a set of independent linear constraints on the motion at a pixel. These initial estimates are then improved by a nonlinear smoothing operation. The results of this algorithm are compared with those obtained by the Horn-Schunck algorithm [10] on a number of image sequences.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)